一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
LOIS2024-85(2025-03)

表情による感情認識を用いた体調推定手法の開発と評価

西山　敦紀[†]　中田　匠哉[†]　陳　思楠[†]　佐伯　幸郎[††]　中村　匡秀[†,†††]

† 神戸大学　〒657-8501 神戸市灘区六甲台町 1-1
†† 高知工科大学　〒782-8502 高知県香美氏土佐山田町宮ノ口 185
††† 理化学研究所・革新知能統合センター　〒103-0027 東京都中央区日本橋 1-4-1
E-mail: †2174391t@stu.kobe-u.ac.jp, ††tnakata@es4.eedept.kobe-u.ac.jp, †††chensinan@gold.kobe-u.ac.jp,
††††saiki.sachio@kochi-tech.ac.jp, †††††masa-n@cmds.kobe-u.ac.jp

あらまし　近年，高齢者の人口は増加しており，一人暮らしの独居高齢者の人口と割合も増加傾向にある．一人暮らしの高齢者は社会的に孤立しやすく，病気やけがの発見が遅れると孤独死につながる可能性が高い．そのような高齢者の日々の体調を把握することは大切であるが，既存の安否確認や見守りサービスでは高齢者が自ら体調不良を伝えるしかないという問題点が存在する．本研究では体調別の演技をした表情を撮影した画像の感情推論結果に主成分分析を適用し，得られた主成分得点を特徴量に機械学習で分類することで体調分類の精度が上がるか評価する．また体調分類モデルの構築を個人ごとに行う必要があるかを検討する．評価した結果，感情推論結果に個人ごとに主成分分析を適用することで精度が上がった体調分類モデルを構築できる．

キーワード　高齢者見守り，表情，感情認識，体調，機械学習

Development and Evaluation of Estimating Health Condition Using Facial Emotion Recognition

Atsunori NISHIYAMA[†], Takuya NAKATA[†], Sinan CHEN[†], Sachio SAIKI[††], and Masahide NAKAMURA[†,†††]

† Kobe University　Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo 657-8501 Japan
†† Kochi University of Technology　185 Miyanokuchi, Tosayamada, Kami City, Kochi 785-8502, JAPAN
††† Riken AIP　1-4-1 Nihon-bashi, Chuo-ku, Tokyo 103-0027 Japan
E-mail: †2174391t@stu.kobe-u.ac.jp, ††tnakata@es4.eedept.kobe-u.ac.jp, †††chensinan@gold.kobe-u.ac.jp,
††††saiki.sachio@kochi-tech.ac.jp, †††††masa-n@cmds.kobe-u.ac.jp

Abstract　In recent years, the population of the elderly has been increasing, and the population and percentage of the elderly living alone are also on the rise. Elderly people who live alone are likely to be socially isolated, and if the detection of illness or injury is delayed, it is likely to lead to a solitary death. It is important to understand the daily health condition of such elderly people, but existing safety confirmation and monitoring services have the problem that elderly people have no choice but to report their health condition by themselves. In this study, we apply principal component analysis to the emotion inference results of images taken of facial expressions of people acting in different health conditions, and evaluate whether the accuracy of health condition classification can be improved by classifying the obtained principal component scores into features using machine learning. We also examine whether it is necessary to construct health condition classification model for each individual. As a result of the evaluation, it is possible to construct health condition classification model with improved accuracy by applying principal component analysis to the emotion inference results for each individual.

Key words　watching over the elderly, facial expressions, facial emotion recognition, health condition, machine learning

This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere.
Copyright ©2025 by IEICE

1. はじめに

近年，高齢者の人口は増加しており，その中でも一人暮らしで生活を送る独居高齢者の割合も増加している [1]．そうした中で高齢者自身が孤立死を身近に感じており，実際に自宅で死亡した一人暮らしの高齢者は多くおり [2]，孤立死と考えられる事例が数多くある．先行研究ではこのような背景から高齢者とコミュニケーションを取ることや見守りを行う，対話エージェントの声掛けによる日常的な体調確認を行い支援につなげることを長期的なゴールとして定め，表情から体調を推定する手法を検討している．しかし，この手法には以下の問題点がある．

P1: 特徴量として happy しか使っていない

P2: 被験者ごとに分類モデル構築をする意義が明確でない

このような課題があり以下のリサーチクエスチョンを設定した．

RQ1: happy 以外の感情値を活用することで体調分類精度を上げられるか

RQ2: 被験者全員共通の高精度なモデルを構築できるか

本研究では特徴量とモデル構築の人数を比較することで体調分類モデルの精度を上げることを目的とする．キーアイデアは七つの感情値に主成分分析を適用すること，そして被験者全員分のデータを用いて主成分分析とモデル構築を行い，被験者ごとに構築した場合と比較することである．目的を達成するために以下の手法を提案する．

A1：七つの感情値に対して主成分分析をする．

A2：主成分分析結果から特徴量を選択する．

A3：個別モデルと全体モデルに分けて機械学習モデルを構築する．

A4：happy のみか主成分分析結果を用いた個別モデルの比較をする．

A5：個別モデルと全体モデルの比較をする．

この手法に沿って評価実験を行う．これにより特徴量には七つの感情値に主成分分析を適用して第 1 主成分と第 2 主成分を用いて，機械学習モデルの構築には被験者ごとに個別にする方が高精度であることが分かった．

2. 準　　備

2.1 独居高齢者の増加

近年，日本の高齢者人口は増加しており令和 6 年度版高齢社会白書 [1] によると，総人口が減少する中で 65 歳以上の者が増加することにより高齢化率は上昇を続け，令和 19 年に 33.3 % となり国民の 3 人に 1 人が 65 歳以上の者となると見込まれている．また 65 歳以上の一人暮らしの者についても 65 歳以上の男女それぞれの人口に占める割合が，令和 2 年には男性 15.0 %，女性 22.1 %となり令和 32 年には男性 26.1 %，女性 29.3 %となると見込まれており独居高齢者も増加している．

独居高齢者については人と話をする頻度について「毎日」と回答した人の割合がひとり暮らし以外の人の半分以下という調査もあり，独居高齢者は高齢者の中でも特に孤立しているという状況が窺える [1]．65 歳以上の者の孤立死に対する意識調査においては，孤立死を身近に感じるかという問いに対して「とても感じる」又は「まあ感じる」と回答したものは 48.7 %である [1] とあり，高齢者自身も孤立死を身近に感じている．実際に，令和 6 年度の上半期に自宅において死亡した一人暮らしの者のうち 65 歳以上は 28330 人に上り [2]，孤立死と考えられる事例が多数発生している．

2.2 face_classification

face_classification [3] は画像から顔を認識し，性別や表情に基づく感情を分類する機械学習モデルを提供している．顔画像の特徴抽出と感情分類には畳み込みニューラルネットワーク (CNN) が使用されており，ソフトマックス活性化関数を適用しているため各感情確度が 0.0〜1.0 に，そしてその総和が 1.0 となるように出力される．データセットには FER-2013 が用いられている．このデータセットは 35887 枚のグレースケール画像が含まれたものであり，各画像が angry, disgust, fear, happy, sad, surprise, neutral のいずれかのクラスに属する．face_classification の感情認識に関するモデルはこのデータセットで 66% の感情分類精度がある [4].

2.3 先行研究：表情による感情認識を用いた体調推定手法の検討

先行研究として，高齢者の体調を把握して声掛けを行うサービスにつなげるために表情から体調を推定できるかを調べ，表情による感情認識を用いた体調推定手法を検討している [5]．高齢者と日々コミュニケーションのと見守りを行うことに注目し，訪問による見守りサービスの人材確保の難しさ，デジタルサービスの高齢者や支援者への負担増加など，また対話エージェントを用いた在宅高齢者の見守りサービスの研究では非言語的な情報を活用できていないという課題を挙げている．そのような課題を受けて長期的なゴールとして対話エージェントの声掛けからテキストと非言語的な情報である表情と声色を取得し，体調を分析してスコアとして算出して日々保存し，スコアが悪ければ詳細な内容を聞くことで支援につなげることを挙げ，その中でも表情から体調を推定できるか調べることをスコープとしている．

体調推定手法では体調を「良い」「普通」「悪い」の 3 段階に分け，事前に体調に応じた表情演技をした動画クリップを撮影する．撮影した動画クリップからフレームを抽出し，既存の感情認識の学習済みモデルを適用して七つの感情クラスの感情値を抽出する．抽出した感情値の中から happy に着目し，この値を特徴量に 3 段階の体調を正解データとして機械学習分類モデルを構築する．構築した分類モデルを用いてテストデータに用いる各フレームの体調を推定し，動画クリップごとに多数決を取る．この手法においては機械学習モデルの構築と体調の推定は個人ごとに行い，複数人のデータを用いて機械学習モデルを構築して体調を推定するということはしていない．

この手法によって体調を推定することができるかを調べるため，体調ごとの表情演技を撮影した動画クリップを収集して機械学習分類モデルを構築して体調を推定する実験を行ったところ，体調を推定できることが示された．

2.4 先行研究の問題点

先行研究では体調を推定する機械学習モデルの構築のプロセ

スにいくつかの課題が残った．

P1:　特徴量として happy しか使っていない

既存の感情認識の学習済みモデルを適用することで動画クリップから抽出した各フレームから angry, disgust, fear, happy, sad, surprise, neutral の感情値が抽出される．そしてその中から特徴量として happy のみに着目して体調推定手法を検討した．しかし抽出できた感情値が七つあるのに対して体調推定手法には happy の一つしか用いておらず，特徴量を増やすことによって体調推定の精度が上がり，より正確に体調の確認を行える可能性がある．

P2:　被験者ごとに分類モデル構築をする意義が明確でない

この実験では体調推定手法として被験者ごとに三つの体調分類の演技をした動画クリップを作成して，そこから抽出したフレームをデータとして機械学習モデルの構築と体調推定を行っている．しかしこの手法では被験者ごとの目の開く大きさや口の開く大きさといった顔の特徴点から計算できる数値を用いておらず，被験者ごとではなく汎用的な感情認識の学習済みモデルを表情に適用した結果の数値を用いている．そのため汎用的な学習済みモデルから得られた数値を用いて機械学習モデルを構築するのであれば，被験者ごとに分ける意義が明確でない．

これら二つの課題を受けて，本研究では二つのリサーチクエスチョンを設定する．

RQ1:　happy 以外の感情値を活用することで体調分類精度を上げられるか

RQ2:　被験者全員共通の高精度なモデルを構築できるか

3. 提 案 手 法

3.1　目的とキーアイデア

本研究の目的は特徴量とモデル構築の人数を比較することで体調分類モデルの精度を上げることである．キーアイデアは七つの感情値に主成分分析を適用すること，そして被験者全員分のデータを用いて主成分分析とモデル構築を行い，被験者ごとに構築した場合と比較することである．目的を達成するため，リサーチクエスチョンに対するアプローチとして以下の五つを提案する．

A1：七つの感情値に対する主成分分析

A2：主成分分析結果から特徴量を選択

A3：個別モデルと全体モデルに分けて機械学習モデルを構築

A4：happy のみか主成分分析結果を用いた個別モデルの比較

A5：個別モデルと全体モデルの比較

3.2　A1: 七つの感情値に対して主成分分析

A1 では機械学習モデルを構築する際の特徴量を作成するために次元削減を行う．表情から感情値を抽出する既存の感情認識の学習済みモデルを適用すると，七つの感情に対してそれぞれ感情値を抽出することができる．しかし七つの感情値をそのままデータとして理解するには 7 次元のデータを理解することになり難しく，そのため次元削減を行いより少ない次元にすることでデータを理解しやすくする．

具体的には七つの感情値を変数にして各主成分の主成分負荷量から主成分得点を算出し寄与率が最も大きい第 1 主成分から

	angry	disgust	fear	happy	sad	surprise	neutral
0	0.097735	0.022408	0.053805	0.291796	0.095147	0.017580	0.421530
1	0.100365	0.018120	0.052083	0.271766	0.112144	0.012496	0.433026
2	0.152202	0.030711	0.057717	0.184227	0.130191	0.032201	0.412751
...	...	...	...	...	...	...	...

主成分分析

	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分	第6主成分	第7主成分
0	-0.021415	-0.070082	0.020271	-0.067434	-0.017918	0.033559	4.879338E-09
1	-0.043716	-0.063761	0.017446	-0.076112	-0.022318	0.048938	-9.474655E-09
2	-0.085417	0.032666	0.024199	-0.097194	-0.014254	0.057504	5.031685E-08
...	...	...	...	...	...	...	...

図 1　七つの感情値データと主成分分析結果の例

第 7 主成分までに変換する．図 1 は七つの感情値データと主成分分析をした結果の例であり，angry, disgust, fear, happy, sad, surprise, neutral の感情値とそれを基に主成分分析をして得られる第 1 主成分から第 7 主成分がある．後のアプローチで被験者ごとに個別に機械学習モデルを構築する場合と全員共通のモデルを構築する場合があるため，それぞれわけて主成分分析を行う必要がある．被験者ごとに個別に主成分分析を行う際には被験者の三つの動画クリップから得た全てのフレームを用いて主成分分析を行い，全員共通で主成分分析を行う際には被験者全員の動画クリップから得た全てのフレームを用いて 被験者人数 × 3 の動画クリップを主成分分析に用いる．

3.3　A2: 主成分分析結果から特徴量を選択

A2 では主成分分析結果から機械学習モデルを構築する際の特徴量を選択する．A1 の変換だけでは 7 次元のデータを用いることになりデータを理解する難易度が変わらない．主成分分析により得られた主成分にはそれぞれ寄与率があり，その主成分がデータの何 % を説明できているのか主成分の重要度を理解できる．選択した主成分で十分に元データを説明するために，第 1 主成分と第 2 主成分を選択し，その累積寄与率を確認する．図 1 においては下の表を見て左から二つを選択することになる．

3.4　A3: 個別モデルと全体モデルに分けて機械学習モデルを構築

A3 では被験者ごとの個別モデルと全員共通の全体モデルに分けて機械学習モデルを構築する．A2 により七つの感情値をすべて用いた新たな特徴量として PCA 特徴量を得ている．次に行うのは本研究の二つの RQ に答える精度の比較をするために，被験者ごとの個別のモデルを構築すること，そして被験者全員に共通する全体モデルを構築することである．RQ1 では happy 以外の感情値を活用することで体調分類精度を上げられるかを調べるために特徴量の違う個別モデルを用意する必要があり，RQ2 では被験者全員共通の高精度なモデルを構築できるかを調べるために特徴量をそろえた個別モデルと全体モデルを用意する必要がある．

具体的には happy 特徴量の個別モデル，PCA 特徴量の個別モデル，happy 特徴量の全体モデル，PCA 特徴量の全体モデルの 4 種類を構築する．これらを構築することにより RQ に答える比較を行うための 4 種類のモデルが準備される．

3.5　A4:happy のみか主成分分析結果を用いた個別モデルの比較

A4 では happy 特徴量の個別モデルと PCA 特徴量の個別モデルを比較する．2 種類のモデルを用いて体調が「良い」「普通」

「悪い」の動画クリップの体調を推定を行い特徴量ごとの比較を行うことで RQ1 に答える．

比較するためにまずは，機械学習モデル構築の際に使わなかったテストデータであるフレームの体調ラベルを個別モデルで推定する．次に推定したラベルを動画クリップごとに集めて多数決を取り，推定される動画クリップの体調ラベルを決定する．そして特徴量ごとの個別モデルと動画クリップの体調推定の精度を比較する．これにより happy 特徴量で十分であるか，あるいは全感情の感情値を用いる方が良いかがわかる．

3.6 A5: 個別モデルと全体モデルの比較

A5 では個別モデルと全体モデルの比較を行う．A3 において 4 種類のモデルを準備しており，これらを用いて動画クリップの体調を推定を行い特徴量をそろえて個別モデルと全体モデルの比較を行うことで RQ2 に答える．

具体的に比較するのは，happy 特徴量の個別モデルと全体モデル，そして PCA 特徴量の個別モデルと全体モデルである．比較するためにまずは，テストデータのフレームの体調を happy 特徴量の全体モデルと PCA 特徴量の全体モデルで推定する．次に happy 特徴量と PCA 特徴量の全体モデルで推定したラベルを動画クリップごとに集めて多数決を取り，推定される動画クリップの体調ラベルをそれぞれのモデルで決定する．最後に happy 特徴量の個別モデルと全体モデルを，そして PCA 特徴量の個別モデルと全体モデルの動画クリップの体調推定の精度を比較する．比較して精度の良し悪しを見ることにより，機械学習モデルを被験者ごとに個別にする意義があるのか，あるいは全員共通のモデルにすることでモデル構築の表情データの数の用意を容易にすることができるかがわかる．

4. 評 価 実 験

RQ を検証するために実験を行った．

4.1 評価実験の概要

RQ1 に対する実験を評価実験 1，RQ2 に対する実験を評価実験 2 とした．実験のデータセットは先行研究 [5] と同じものを用いており，特徴量やモデル構築の方法を変えることにより RQ1 と RQ2 の結論を出す．被験者は研究者とその家族及び我々の研究室のメンバーに協力していただき，10 代女性 1 名，20 代男性 9 名，50 代男性 1 名女性 1 名の 12 名を選出した．既存の感情認識の学習済みモデルには face_classification の感情認識を用いており，データセットは動画クリップ内の各フレームに感情認識を適用して抽出した感情値ベクトルを保存したデータフレームであった．

体調を推定する機械学習分類モデルの構築と評価には Python を使用し，ライブラリは numpy, pandas, matplotlib, seaborn, scikit-learn を用いた．分類モデルには決定木を採用し，scikit-learn の DecisionTreeClassifier を用いた．

4.2 評価実験 1 の方法

まずは七つの感情クラスを理解しやすく次元削減するために，被験者ごとに主成分分析を行った．被験者ごとに「良い」「普通」「悪い」の三つの動画クリップから抽出した感情値ベクトルのデータフレームを一つにし，python の scikit-learn ライブラリを用いてデータセットに主成分分析 (PCA) を適用した．主成分分析を適用して第 1 主成分と第 2 主成分で約 90% の累積寄与率であることを確認し，その二つの主成分の主成分負荷量をグラフに表した．

次に happy 特徴量と PCA 特徴量を用いて被験者ごとに個別の機械学習分類モデルを構築した．訓練データとテストデータは scikit-learn の train_test_split のパラメータ，test_size を 0.5 に設定して被験者ごとにデータセットを半分に分けた．またパラメータとして stratify=y も設定し，各体調ラベルが訓練，テストデータで均等になるようにした．個別の分類モデルを構築する際のハイパーパラメータをグリッドサーチと 6 分割交差検証によって，木の最大の深さである max_depth と決定木を分類する際の最小のサンプルサイズである min_samples_split の二つを決定した．グリッドサーチはパラメータのすべての組み合わせを試してみる方法であり，scikit-learn の GridSerarchCV ではパラメータの組合せに対して交差検証が実行され，最適なパラメータ設定をサーチする [6]．max_depth は 1, 2, 3, 4, 5 の中から，min_samples_split は 2, 10, 20, 30 の中から設定した．また DecisionTreeClassifier のパラメータに class_weight='balanced' を設定し，不均衡データに考慮するようにした．

個別モデルを構築した後はテストデータについて，個別モデルを用いて happy 特徴量と PCA 特徴量からフレームの体調ラベルを推定した．個別モデルの評価はマクロ平均 F 値で行い，各被験者のマクロ平均 F 値と被験者全員の平均をグラフに表した．マクロ平均 F 値とは各体調の適合率と再現率から計算した F 値の平均である．

次に同じ動画クリップから抽出されたテストデータの推定した体調を集めて多数決を取った．そして被験者の三つの体調の動画クリップがどの体調だと推定されたか被験者全員分を一つの混同行列に表した．

4.3 評価実験 1 の結果

まず累積寄与率を確認し，被験者 E の 86%，被験者 H の 89% 以外は 90% を超えており，全員が約 90% になっていることを確認した．個別の第 1 主成分と第 2 主成分の主成分負荷量を図 2 に示す．happy 特徴量の個別モデルと PCA 特徴量の個別モデルのハイパーパラメータをそれぞれ表 1 と表 2 に表す．happy 特徴量の個別モデルと PCA 特徴量の個別モデルのマクロ平均 F 値の棒グラフとその平均の点線を図 3 と図 4 に示す．そして happy 特徴量の個別モデルと PCA 特徴量の個別モデルでそれぞれ体調推定を行った結果を表す混同行列を図 5 と図 6 に示す．

初めに被験者ごとの主成分負荷量について，被験者 E を除き happy のみが突出して大きい正の負荷量を持ち neutral 以外がそれに対して余り大きい負荷量を持たない主成分があり，被験者 E についても happy が最も大きい負荷量となる第 2 主成分が見られた．またそれ以外のもう一つの主成分では angry, disgust, fear, sad が大きい正の負荷量となるような主成分を持つことがわかった．

次に個別モデルのマクロ平均 F 値は特徴量が happy 特徴量の場合，最も大きい被験者で 0.89，小さい被験者で 0.49，平均で約 0.69 となり，PCA 特徴量の場合は最も大きい被験者で 0.95,

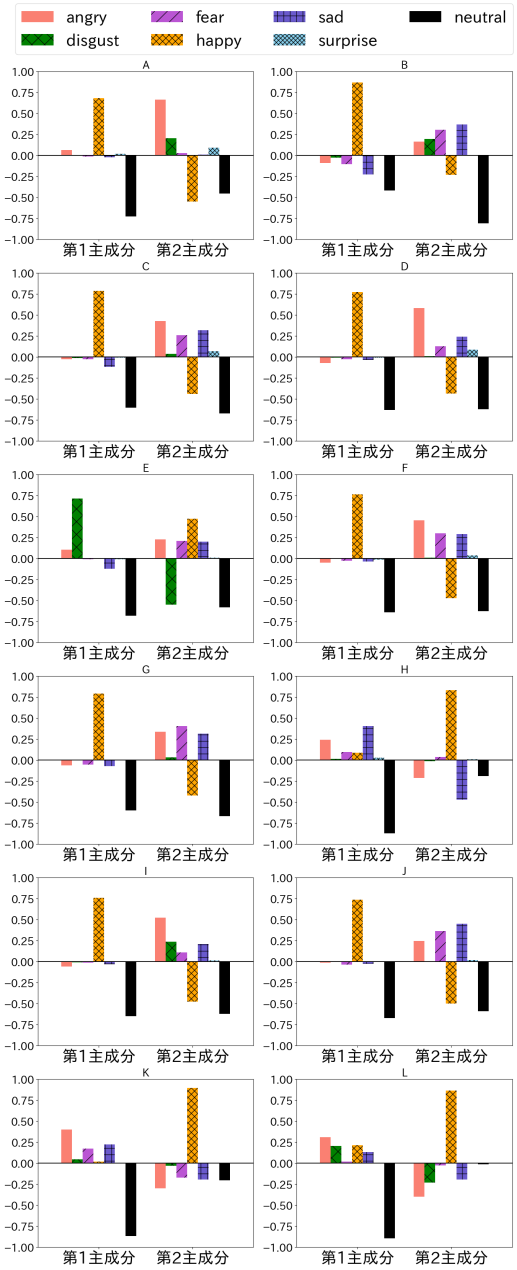

図 2　個別の第 1 主成分と第 2 主成分の主成分負荷量

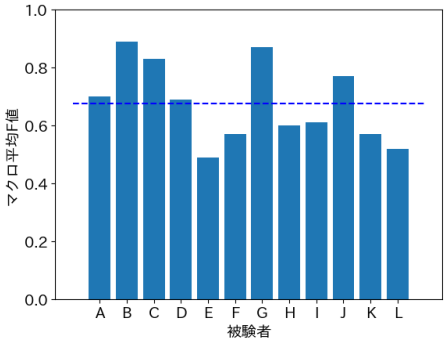

図 3　happy 特徴量の個別モデルのマクロ平均 F 値とその平均

小さい被験者で 0.62，平均で約 0.83 となりどれも PCA 特徴量の方が高くなった．そして動画クリップの体調推定結果では happy 特徴量はパラメータを変更することで全体のバランス良く改善がされているが，PCA 特徴量のほうが「悪い」ラベルの一つを除き推定できているという結果になった．

表 1　happy 特徴量の個別モデルのハイパーパラメータ

	max_depth	min_samples_split
被験者 A	5	30
被験者 B	2	2
被験者 C	4	30
被験者 D	3	2
被験者 E	1	2
被験者 F	4	2
被験者 G	4	20
被験者 H	2	2
被験者 I	2	2
被験者 J	5	30
被験者 K	5	2
被験者 L	4	20

表 2　PCA 特徴量の個別モデルのハイパーパラメータ

	max_depth	min_samples_split
被験者 A	2	2
被験者 B	5	10
被験者 C	5	10
被験者 D	5	2
被験者 E	5	10
被験者 F	3	2
被験者 G	4	30
被験者 H	5	10
被験者 I	4	2
被験者 J	5	2
被験者 K	5	2
被験者 L	5	20

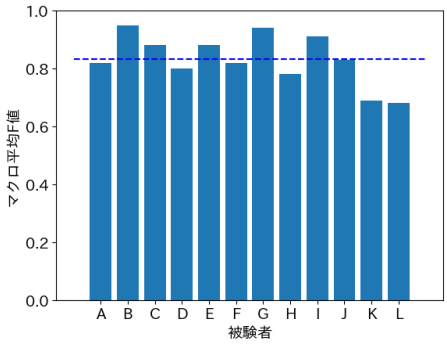

図 4　PCA 特徴量の個別モデルのマクロ平均 F 値とその平均

4.4　評価実験 2 の方法

まずは被験者全員共通の全体モデルを構築するために，被験者全員のデータを使って主成分分析を行った．被験者全員の動画クリップから抽出した感情値ベクトルのデータフレームを一つにし，python の scikit-learn ライブラリを用いて主成分分析を適用した．主成分分析を適用して第 1 主成分と第 2 主成分で約 90% の累積寄与率であることを確認し，その二つの主成分の主成分負荷量をグラフに表した．

次に happy 特徴量と PCA 特徴量を用いてぞれぞれ体調を推

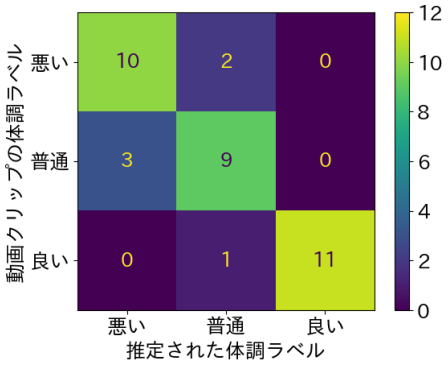

図 5　happy 特徴量の個別モデルでの体調推定結果を表す混同行列

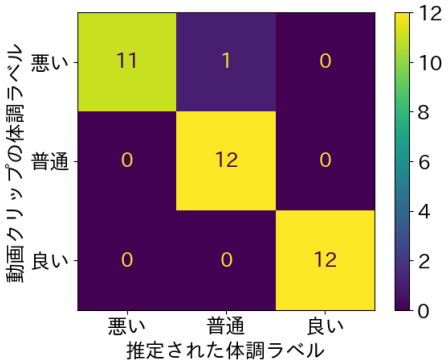

図 6　PCA 特徴量の個別モデルでの体調推定結果を表す混同行列

定する被験者全体共通の機械学習分類モデルを構築した．訓練データとテストデータは全体のデータセットをラベルが均等になるように半分に分けた．全体の分類モデルを構築する際のハイパーパラメータをグリッドサーチと 6 分割交差検証によって，max_depth と min_samples_split の二つを決定した．max_depth は 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 の中から，min_samples_split は 2, 10, 20, 30 の中から設定し，不均衡データに考慮するように全体モデルを構築した．

個別モデルを構築した後はテストデータについて，全体モデルを用いて happy 特徴量と PCA 特徴量の二通りについてフレームの体調ラベルを推定した．フレームを推定する全体モデルの評価はマクロ平均 F 値で行い，全てのテストデータを被験者ごとに分けて評価するために各被験者のマクロ平均 F 値と被験者全員の平均をグラフに表した．

次に同じ動画クリップから抽出されたテストデータの推定した体調を集めて多数決を取った．そして動画クリップがどの体調だと推定されたか混同行列に表した．

マクロ平均 F 値を表したグラフとテストデータを用いた動画クリップの体調推定結果を表した混同行列の二つについて，happy 特徴量の個別モデルと全体モデルを比較し，PCA 特徴量の個別モデルと全体モデルで比較した．また構築した分類モデルが PCA 特徴量を用いてどのように体調を分類しているかを確認するために，x 軸を第 1 主成分に，y 軸を第 2 主成分として被験者ごとにテストデータを散布図に表し，散布図の背景を個別モデル分類される体調ラベルで色分けした．また全体モデルでも同様に分類される体調ラベルで色分けした．

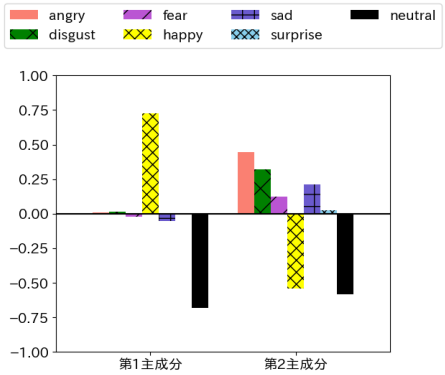

図 7　全員共通の第 1 主成分と第 2 主成分の主成分負荷量

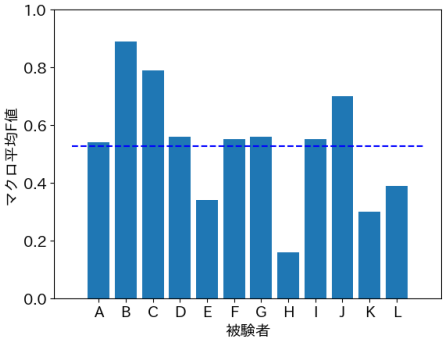

図 8　happy 特徴量の全体モデルのマクロ平均 F 値とその平均

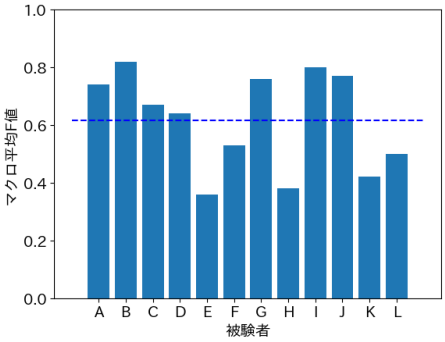

図 9　PCA 特徴量の全体モデルのマクロ平均 F 値とその平均

4.5　評価実験 2 の結果

まず累積寄与率が 93% であることを確認した．全体で主成分分析を行った結果の主成分負荷量を図 7 に示す．happy 特徴量の全体モデルのハイパーパラメータは max_depth が 2，min_samples_split が 2，PCA 特徴量では max_depth が 5，min_samples_split が 2 になった．また happy 特徴量の全体モデルと PCA 特徴量の全体モデルのマクロ平均 F 値の棒グラフとその平均の点線をそれぞれ図 8 と図 9 に示す．そして happy 特徴量の全体モデルと PCA 特徴量の全体モデルで体調推定を行った結果を表す混同行列をそれぞれ図 10 と図 11 に示す．各被験者の PCA 特徴量でのテストデータのプロットと決定木による分類モデルを背景に表した図を個別モデルと全体モデルでそれぞれ図 12 と図 13 に示す．

まず全体主成分分析を行った結果の主成分負荷量について，第 1 主成分が happy のみが大きな正の負荷量を持つ主成分となり，

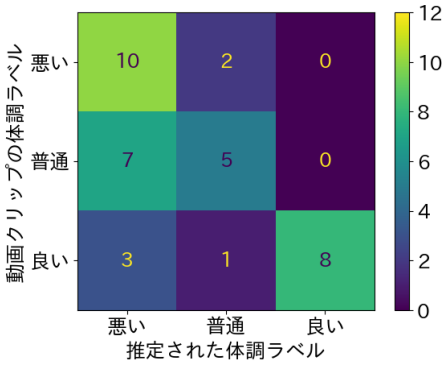

図 10　happy 特徴量の全体モデルでの体調推定結果を表す混同行列

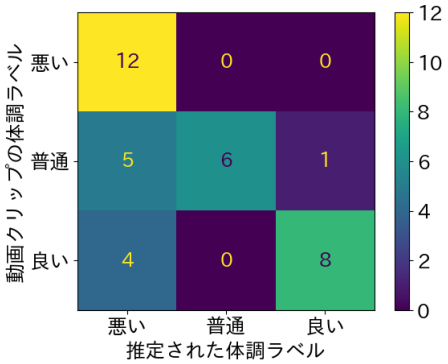

図 11　PCA 特徴量の全体モデルでの体調推定結果を表す混同行列

第 2 主成分が angry, disgust, fear, sad, そしてわずかに surprise が正の負荷量となった．

次に happy 特徴量の全体モデルで各被験者のテストデータの体調ラベルを推定した結果，マクロ平均 F 値は最も大きい被験者で 0.89，小さい被験者で 0.16，平均で約 0.53 となり，PCA 特徴量では最も大きい被験者で 0.83，小さい被験者で 0.36，平均で約 0.62 となった．happy 特徴量の個別モデルと全体モデルを比較すると個別モデルのほうが最大値，最小値，平均値がどれも大きいという結果になり，PCA 特徴量についても同様だった．そして動画クリップの体調推定結果では happy 特徴量では「悪い」ラベルの再現率が変わらないが「良い」「普通」ラベルでは個別のほうがよかった．PCA 特徴量では全体モデルは「普通」ラベルの再現率が 0.50 と happy のみの場合より改善して全体としてマクロ平均 F 値が 0.72 という値になったが，こちらも個別モデルのほうがよいという結果だった．

最後に散布図について，被験者ごとに主成分分析と決定木分類を行うとテストデータが体調ごとに分散し，決定木分類によってその集まりをとらえられた．被験者全員共通で主成分分析と決定木分類を行うと個別の場合よりも主成分によってテストデータが分散せず別の体調のテストデータ同士が重なり，決定木分類によってその重なった箇所を誤って分類された．

4.6　考　　察

評価実験 1 と評価実験 2 を通して，まず七つの感情クラスに主成分分析を行った結果について，第 1 主成分と第 2 主成分のどちらかが happy のみが大きな正の負荷量を持つ主成分となることから，特徴量として happy の値を用いるという選択は特徴量を一つ選んでモデルを構築する場合は適した選択であったと考える．

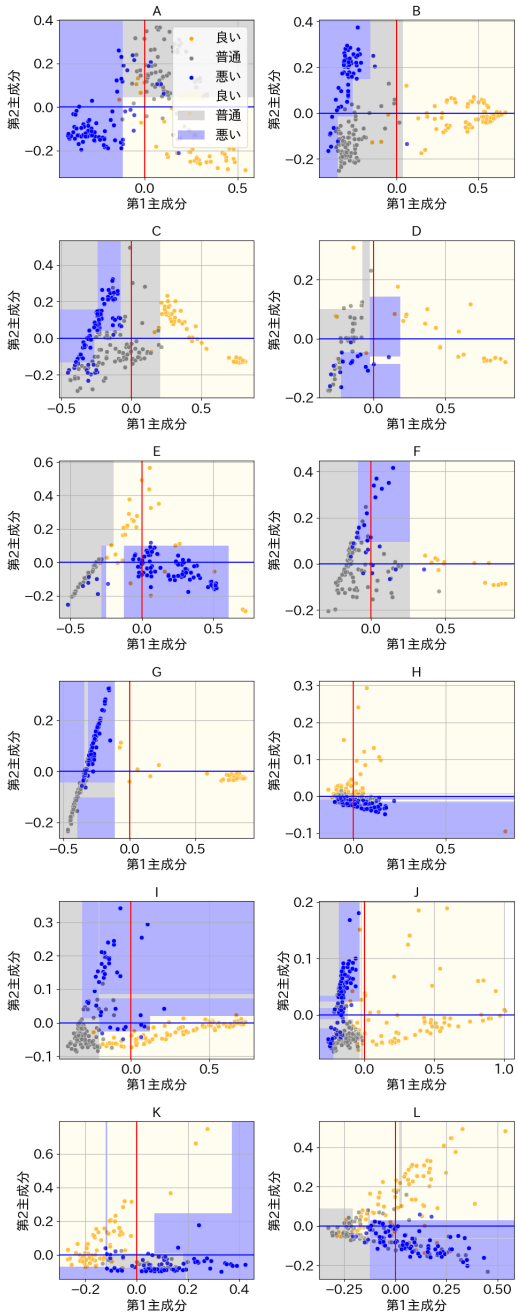

図 12　被験者ごとのテストデータの散布図と個別モデルによる分類

次に評価実験 1 の特徴量の違いによる個別モデルの比較では，第 1 主成分と第 2 主成分を用いることで体調推定したフレームの多数決を取る前の F 値についても happy のみより高く，平均して 0.9 に達していないが多数決を取ることで動画クリップの体調を推定できている．よって RQ1 としては表情による感情認識で抽出できる感情クラスを全て特徴量に用いる必要がある．

また評価実験 2 の個別モデルと全体モデルの比較では，特徴量として happy のみを用いた個別モデルでは「良い」について良く正解できておりこれは happy という感情によってより笑顔が顕著なものが分類できたためだと先行研究では考察しているが，全体モデルでは再現率が下がっている．一方で「悪い」ラ

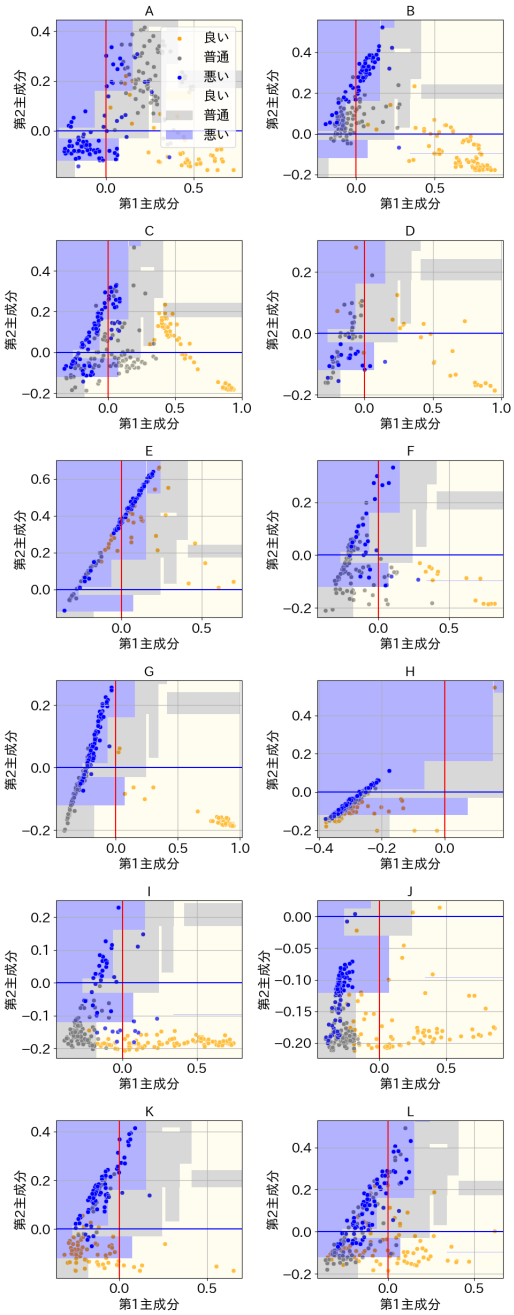

図 13　被験者ごとのテストデータの散布図と全体モデルによる分類

ベルの再現率は全体モデルも変わらなかったが「良い」「普通」の動画クリップも悪いと推定するケースが多く，「普通」では再現率が 0.5 を下回っている．全体モデルは「悪い」ラベルに推定が偏ったモデルになっているため表情から体調を推定した結果を受けて高齢者の体調を伺うために用いると考えると体調不良を見逃して悪化させてしまうことが少ないモデルではあるが，個別モデルのほうが正確に体調に合わせた声掛けにつなげることができると考える．PCA 特徴量の全体モデルも「悪い」ラベルを見逃さないようなモデルとなっており悪い場合の再現率が 1.0 あるが，残りのラベルで個別モデルに劣るため，全体モデルによって高精度なモデルが構築できるとはいえない．以上より RQ2 としては全体モデルでは個別モデルほど高精度にはならない．最後に主成分分析後のテストデータの散布図と決定木による分類モデルについて，全体モデルでは主成分分析によって被験者によってはテストデータをうまく体調ごとに分散させることができなかったため，決定木分類が個別モデル程高精度にはならなかったと考える．

実験の結果，表情による感情認識を用いた体調推定手法として特徴量には七つの感情クラスに主成分分析を適用して第 1 主成分と第 2 主成分を特徴量として用いて，機械学習分類モデルの構築には被験者ごとに個別にすることで表情から体調を推定する分類モデルを構築できることがわかった．

5. ま と め

本研究では表情による感情認識を用いた体調推定手法の開発と評価を行った．機械学習モデル構築の際の特徴量として happy 以外の感情値を活用することで体調分類精度を上げられるか，被験者個別ではなく全員共通の高精度なモデルを構築できるかという二つの RQ を立て，特徴量とモデル構築の人数を比較することで体調分類モデルの精度を上げることを目的として評価実験を行った．アプローチとしては，「A1：七つの感情値に対する主成分分析」，「A2：主成分分析結果から特徴量を選択」，「A3：個別モデルと全体モデルに分けて機械学習モデルを構築」，「A4：happy のみか主成分分析結果を用いた個別モデルの比較」，「A5：個別モデルと全体モデルの比較」の五つである．この提案手法に基づきモデルの作成と評価を行った．

その結果，特徴量には七つの感情値に主成分分析を適用して第 1 主成分と第 2 主成分を用いて，機械学習モデルの構築には被験者ごとに個別にする方が高精度であることが示された．今後の課題として，推定した体調から対話エージェントとの対話により深掘りして詳細を聞き出すサービスを開発すること，実験以外の状況で動画クリップを収集する方法を検討すること，会話のテキストや他の非言語的な情報である音声でも体調を分類できるか実験することが挙げられる．

謝辞　本研究の一部は JSPS 科研費 JP20H05706, JP22H03699, JP22K19653, JP23H03401, JP23H03694, JP23K17006，および，立石科学技術振興財団の研究助成を受けて行われている．

文　　献

[1] 内閣府，“令和 6 年度版高齢社会白書,” https://www8.cao.go.jp/kourei/whitepaper/w-2024/zenbun/slash06pdf_index.html. accessed October 17, 2024.

[2] 警察庁刑事局捜査第一課，“警察取扱死体のうち、自宅において死亡した一人暮らしの者～令和 6 年上半期（1 6 月）暫定値～,” https://www.npa.go.jp/publications/statistics/shitai/hitorigurashi/240827_kenshi2.pdf. accessed October 17, 2024.

[3] “face_classification,” https://github.com/oarriaga/face_classification. accessed October 17, 2024.

[4] M.V. Octavio Arriaga, Paul G. Plöger, “Real-time convolutional neutral networks for emotion and gender classification,” accessed October 17, 2024, ••.

[5] 西山敦紀，中田匠哉，陳思楠，佐伯幸郎，中村匡秀，“表情による感情認識を用いた体調推定手法の検討,” 電子情報通信学会技術研究報告，第 124 巻，pp.56–64，Nov. 2024.

[6] A.C. Müller and S. Guido, Python ではじめる機械学習—scikit-learn で学ぶ特徴量エンジニアリングと機械学習の基礎，株式会社オライリー・ジャパン，2017.